**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/31, 15/74, A61K 39/04, C07K 14/35, C12Q 1/68 — A2

(11) International Publication Number: **WO 95/17511**

(43) International Publication Date: 29 June 1995 (29.06.95)

(21) International Application Number: PCT/US94/14912

(22) International Filing Date: 23 December 1994 (23.12.94)

(30) Priority Data:

| | | |
|---|---|---|
| 250584 | 23 December 1993 (23.12.93) | NZ |
| 08/201,880 | 24 February 1994 (24.02.94) | US |
| 08/265,579 | 24 June 1994 (24.06.94) | US |
| 08/292,695 | 18 August 1994 (18.08.94) | US |

(60) Parent Application or Grant
(63) Related by Continuation
US 08/292,695 (CIP)
Filed on 18 August 1994 (18.08.94)

(71) Applicants *(for all designated States except US)*: AGRESEARCH NEW ZEALAND PASTORAL AGRICULTURE RESEARCH INSTITUTE, LTD. [NZ/NZ]; P.O. Box 40063, Ward Street, Upper Hutt (NZ). ALBERT EINSTEIN COLLEGE OF MEDICINE OF YESHIVA UNIVERSITY [US/US]; P.O. Box 40063, Ward Street, Upper Hutt (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: JACOBS, William, R., Jr. [US/US]; 163 Fordham Street, City Island, NY 10464 (US). BLOOM, Barry, R. [US/US]; 61 Summit Drive, Hastings-on-Hudson, NY 10706 (US). COLLINS, Desmond, Michael [NZ/NZ]; 23 Stephens Street, Upper Hutt, Wellington (NZ). DE LISLE, Geoffrey, W. [NZ/NZ]; 41 Vista Crescent, Upper Hutt, Wellington (NZ). PASCOPELLA, Lisa [US/US]; 215 South 7th Street, Hamilton, MT 59840 (US). KAWAKAMI, Rika, Pamela [NZ/NZ]; 1 Church Lane, Point Howard, Wellington (NZ).

(74) Agents: MONROY, Gladys, R. et al.; Morrison & Foerster, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ).

Published
*Without international search report and to be republished upon receipt of that report.*

(54) Title: MYCOBACTERIA VIRULENCE FACTORS AND A METHOD FOR THEIR IDENTIFICATION

(57) Abstract

The present invention provides polynucleotides associated with virulence in mycobacteria, and particularly a fragment of DNA isolated from *M. bovis* that contains a region encoding a putative sigma factor. Also provided are methods for a DNA sequence or sequences associated with virulence determinants in mycobacteria, and particularly in *M. tuberculosis* and *M. bovis*. The invention also provides corresponding polynucleotides associated with avirulence in mycobacteria. In addition, the invention provides a method for producing strains with altered virulence or other properties which can themselves be used to identify and manipulate individual genes.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| AT | Austria | GB | United Kingdom | MR | Mauritania |
| AU | Australia | GE | Georgia | MW | Malawi |
| BB | Barbados | GN | Guinea | NE | Niger |
| BE | Belgium | GR | Greece | NL | Netherlands |
| BF | Burkina Faso | HU | Hungary | NO | Norway |
| BG | Bulgaria | IE | Ireland | NZ | New Zealand |
| BJ | Benin | IT | Italy | PL | Poland |
| BR | Brazil | JP | Japan | PT | Portugal |
| BY | Belarus | KE | Kenya | RO | Romania |
| CA | Canada | KG | Kyrgyzstan | RU | Russian Federation |
| CF | Central African Republic | KP | Democratic People's Republic of Korea | SD | Sudan |
| CG | Congo | | | SE | Sweden |
| CH | Switzerland | KR | Republic of Korea | SI | Slovenia |
| CI | Côte d'Ivoire | KZ | Kazakhstan | SK | Slovakia |
| CM | Cameroon | LI | Liechtenstein | SN | Senegal |
| CN | China | LK | Sri Lanka | TD | Chad |
| CS | Czechoslovakia | LU | Luxembourg | TG | Togo |
| CZ | Czech Republic | LV | Latvia | TJ | Tajikistan |
| DE | Germany | MC | Monaco | TT | Trinidad and Tobago |
| DK | Denmark | MD | Republic of Moldova | UA | Ukraine |
| ES | Spain | MG | Madagascar | US | United States of America |
| FI | Finland | ML | Mali | UZ | Uzbekistan |
| FR | France | MN | Mongolia | VN | Viet Nam |
| GA | Gabon | | | | |

# MYCOBACTERIA VIRULENCE FACTORS AND A METHOD FOR THEIR IDENTIFICATION

## Technical Field

This invention relates to polynucleotide sequence(s) associated with virulence in mycobacteria, methods for isolating such sequence(s), and the use of such sequence(s) in human and animal medical practice. It also relates to polypeptides encoded in the sequences.

## Background Art

The mycobacteria are rod-shaped, acid-fast, aerobic bacilli that do not form spores. Several species of mycobacteria are pathogenic to humans and/or animals, and determining factors associated with their virulence are of prime importance. For example, tuberculosis is a worldwide health problem which causes approximately 3 million deaths each year (17), yet little is known about the molecular basis of tuberculosis pathogenesis. The disease is caused by infection with *Mycobacterium tuberculosis*; tubercle bacilli are inhaled and then ingested by alveolar macrophages. As is the case with most pathogens, infection with *M. tuberculosis* does not always result in disease. The infection is often arrested by a developing cell-mediated immunity (CMI) resulting in the formation of microscopic lesions, or tubercles, in the lung. If CMI does not limit the spread of *M. tuberculosis*, caseous necrosis, bronchial wall erosion, and pulmonary cavitation may occur. The factors that determine whether infection with *M. tuberculosis* results in disease are incompletely understood.

The tuberculosis complex is a group of four mycobacterial species that are so closely related genetically that it has been proposed that they be combined into a single species. Three important members of the complex are *Mycobacterium tuberculosis*, the major cause of human tuberculosis; *Mycobacterium africanum*, a major cause of human tuberculosis in some populations; and *Mycobacterium bovis*, the cause of bovine tuberculosis. None of these mycobacteria is restricted to being pathogenic for a single host species. For example, *M. bovis* causes tuberculosis in a wide range of animals including humans in which it causes a disease that is clinically indistinguishable from that caused by *M. tuberculosis*. Human tuberculosis is a major cause of mortality throughout the world, particularly in less developed countries. It accounts for approximately eight million new cases of clinical disease and three million deaths each year. Bovine tuberculosis, as well as causing a small percentage of these human cases, is a major cause of animal suffering and large economic costs in the animal industries.

Antibiotic treatment of tuberculosis is very expensive and requires prolonged administration of a combination of several antituberculosis drugs. Treatment with single antibiotics is not advisable as tuberculosis organisms can develop resistance to the therapeutic levels of all antibiotics that are effective against them. Strains of *M. tuberculosis* that are resistant to one or more antituberculosis drugs are becoming more frequent and treatment of patients infected with such strains is expensive and difficult. In a small but increasing percentage of human tuberculosis cases the tuberculosis organisms have become resistant to the two most useful antibiotics, isoniazid and rifampicin. Treatment of these patients presents extreme difficulty

and in practice is often unsuccessful. In the current situation there is clearly an urgent need to develop new methods for detecting virulent strains of mycobacteria and to develop tuberculosis therapies.

There is a recognized vaccine for tuberculosis which is an attenuated form of *M. bovis* known as BCG. This is very widely used but it provides incomplete protection. The development of BCG was completed in 1921 but the reason for its avirulence was and has continued to remain unknown (Grange et al., 1983). Methods of attenuating tuberculosis strains to produce a vaccine in a more rational way have been investigated but have not been successful for a variety of reasons (Young, 1993). However, in view of the evidence that dead *M. bovis* BCG was less effective in conferring immunity than live BCG (Block and Segal, 1955), there exists a need for attenuated strains of mycobacteria that can be used in the preparation of vaccines.

A variety of compounds have been proposed as virulence factors for tuberculosis but, despite numerous investigations, good evidence to support these proposals is lacking. Nevertheless, the discovery of a virulence factor or factors for tuberculosis is still regarded as important and is a very active area of current research. This is because such a discovery would not only enable the possible development of a new generation of tuberculosis vaccines but might also provide a target for the design or discovery of new or improved anti-tuberculosis drugs or therapies.

The ability to transfer and express recombinant DNA among the mycobacteria, first demonstrated in 1987 (Jacobs et al.), enables the usage of molecular genetics to elucidate pathogenic mechanisms. But, the present lack of evidence of homologous recombination in the pathogenic mycobacteria

has prevented the application of allele exchange systems (Kalpana et al.) for the analysis.

One of the first examples of *in vivo* selection for virulent bacteria was demonstrated by the classic work of Griffith et al. in 1928. Griffith using pneumococci observed that, as a result of genetic exchange, virulent, capsulated pneumococci were recovered from mice infected with a mixture of live attenuated, non-capsulated pneumococci and heat-killed capsulated pneumococci (Griffith, 1928). However, similar systems have not been demonstrated in mycobacteria.

Bacterial RNA polymerases are composed of a core enzyme with the subunit composition $\alpha_2\beta\beta'$ and one of a variety of sigma factors. Transcription responses to changes in growth conditions are modulated by multiple RNA polymerases having different sigma factors which promote transcription of different classes of promoters. The principal sigma factor plays a central role in bacterial by promoting essential "housekeeping" genes. Genes for alternative sigma factors are present in all bacteria and have been shown to promote specific virulence genes in some pathogens (Fang, 1992; Deretic 1994). However, loss of a virulence phenotype due to mutation in a principal sigma factor has not been reported. *Streptomyces sp.* contain several homologues of principal sigma factors (Buttner, 1990) which are not essential for normal growth but which appear to have a function under certain growth conditions.

References cited in the Background Art

1. Anon (1972) TRUDEAU MYCOBACTERIAL CULTURE COLLECTION (Trudeau Institute Inc., P.O. Box 59, Saranac Lake, New York 12983).

2. Belisle, J.T., Pascopella, L., Inamine, J.M., Brennan, P.J., and W.R. Jacobs, (1991) "Isolation and expression of a gene cluster responsible for biosynthesis of the glycopeptidolipid antigens of *Mycobacterium avium*," *J. Bacteriol.* **173**:6991-6997.

3. Bloom, B.R., Tuckman, M., Kalpana, G.V., and W. R. Jacobs, unpublished data.

4. Boyer, H., and D. Roulland-Dussoin, (1969) "A complementation analysis of the restriction and modification of DNA in *Escherichia coli*," *J. Mol. Biol.* **41**:459-472.

5. Collins, D. M., and G. W. de Lisle, (1984) "DNA restriction endonuclease analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG," *J. Gen. Microbiol.* **130**:1019-1021.

6. Collins, F. M., and M. N. Smith, (1969) "A comparative study of the virulence of *Mycobacterium tuberculosis* measured in mice and guinea pigs," *American Review of Respiratory Disease* **100**:631-639.

7. Collins, D. M., S. K. Erasmuson, D. M. Stephens, G. F. Yates, and G. W. de Lisle, (1993) "DNA fingerprinting of *Mycobacterium bovis* strains by restriction fragment analysis and hybridization with the insertion elements *IS1081* and *IS6110*," *J. Clin. Microbiol.* **31**:1143-1147.

8. Dannenberg, A.M., Jr., (1991) "Delayed type hypersensitivity and cell mediated immunity in the pathogenesis of tuberculosis," *Immunology Today* **12**:228-233.

9. Gallagher, J., and D. M. Horwill, (1977) "A selective oleic acid albumin medium for the cultivation of *Mycobacterium bovis*," *J. Hyg. Camb.* **79**:155-160.

10. Grange, J. M., J. Gibson, T. W. Osborne, C. H. Collins and M. D. Yates, (1983) "What is BCG?" *Tubercle* **64**:129-139.

11. Griffith F., (1928) "Significance of pneumococcal types," *J. Hyg.* **27**:113-159.

12. Grosskinsky, C.M., Jacobs, W.R., Jr., Clark-Curtiss, J.E., and B.R. Bloom, (1989) "Genetic relationships among *Mycobacterium leprae*, *Mycobacterium tuberculosis*, and candidate leprosy vaccine strains determined by DNA hybridization: Identification of an *M. leprae*-specific repetitive sequence," *Infect. Immun.* **57**:1535,1541.

13. Jacobs, W.R., Barrett, J.F., Clark-Curtiss, J.E., and R. Curtiss III, (1986) "*In vivo* repackaging of recombinant cosmid molecules for analysis of *Salmonella typhimurium*, *Streptococcus mutans*, and mycobacterial genomic libraries," *Infect. Immun.* **52**:101-109.

14. Jacobs, W. R., G. V. Kalpana, J. D. Cirillo, L. Pascopella, S. B. Snapper, R. A. Udani, W. Jones, R. G. Barletta and B. R. Bloom, (1991) "Genetic systems for mycobacteria," *Methods Enzymol.* **204**:537-555.

15. Jacobs, W.R., Tuckman, M., and B.R. Bloom, (1987) "Introduction of foreign DNA into mycobacteria using a shuttle plasmid," *Nature* **327**:532-535.

16. Kalpana, G.V., Bloom, B.R., and W.R. Jacobs, (1991) "Insertional mutagenesis and illegitimate recombination in mycobacteria," *Proc. Natl. Acad. Sci. USA* **88**:5433-5437.

17. Kochi, A., (1991) "The global tuberculosis situation and the new control strategy of the World Heath Organization," *Tubercle* **72**:1-12.

18. Lee, M.H., Pascopella, L., Jacobs, W.R., and C.F. Hatfull, (1991) "Site-specific integration of mycobacteriophage L5: Integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Clamette-Guerin," *Proc. Natl. Acad. Sci. U.S.A.* **88**:3111-3115.

19. Mackaness, G.B., Smith, N., and A.Q. Wells, (1954) "The growth of intracellular tubercle bacilli in relation to their virulence," *Am. Rev. Tuberc.* **69**:479-494.

20. North, R.J., and A.A. Izzo, (1993) "Mycobacterial virulence: Virulent strains of *Mycobacterium tuberculosis* have faster *in vivo* doubling times and are better equipped to resist growth inhibiting functions of the macrophages in the presence and absence of specific immunity," *J. Exp. Med.* **177**:1723-1734.

21. Oatway, W.H., Jr., and W. Steenken, Jr., (1936) "The pathogenesis and fate of tubercle produced by dissociated variants of tubercle bacilli," *J. Inf. Dis.* **59**:306-325.

22. Pascopella, L., F. M. Collins, J. M. Martin, M. H. Lee, G. F. Hatfull, B. R. Bloom and W. R. Jacobs, "*In vivo* complementation in *Mycobacterium tuberculosis* to identify a genomic fragment associated with virulence," Infection and Immunity 62: 1313-1319.

23. Pierce, C.H., Dubos, R.J., and W.B. Schaefer, (1953) "Multiplication and survival of tubercle bacilli in the organs of mice," *J. Exp. Med.* **97**:189-206.

24. Sambrook, J., E. F. Fritsch and T. Maniatis, (1989) MOLECULAR CLONING: A LABORATORY MANUAL (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

25. Steenken, W., Jr., and L.U. Gardner, (1946.) "History of H37 strain tubercle bacillus," *Amer. Rev. Tuberc.* **54**:62-66.

26. Steenken, W., Jr., Oatway, W.H., Jr., and S.A. Petroff, (1934) "Biological studies of the tubercle bacillus. III. Dissociation and pathogenicity of the R and S variants of the human tubercle bacillus ($H_{37}$)," *J. Exp. Med.* **60**:515-540.

27. Stover, C.K., de la Cruze, V.F., Fuerst, T.R., Burlein, J.E., Benson L.A., Bennett L.T., Bansal,

G.P., Young, J.F., Lee, M.H., Hatful, G.F., Snapper, S.B., Barletta, R.G., Jacobs, W.R., and B.R. Bloom, (1991) "New use of BCG for recombinant vaccines," *Nature* **351**:456-460.

28. van Soolingen, D., P. W. M. Hermans, P. E. W. de Haas, D. R. Soll and J. D. A. van Embden, (1991) "Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis," *J. Clin. Microbiol.* **29**:2578-2586.

29. Weis, H., (1991) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Supplement 13, 5.3. (Greene Publishing Associates, New York, eds., F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Siedman, J. A. Smith, and K. Struhl).

30. Young, D. B. and S. T. Cole, (1993) "Leprosy, tuberculosis, and the new genetics," *J. Bacteriol.* **175**:1-6.

31. Block, H. and W. Segal, Am. Rev. Tuberc. Pulm. Dis 71:228-248.

32. Fang, C.F. et al., (1992) Proc. Natl. Acad. Sci. USA 89:11978-11982.

33. Deretic., V. et al, (1994) J. Bact. 176:2773-2780.

34. Buttner, M.J., et al., (1990) J. Bact. 172:3367-3378.

Summary of the Invention

The present invention provides isolated and recombinant polynucleotide sequences associated with virulence determinants in members of the genus mycobacteria, particularly those of the tuberculosis complex, and more particularly in *M. tuberculosis* and *M. bovis*. Based upon homology to sigma factors from other

microorganisms, one of the mycobacterial sequences associated with virulence encodes a putative sigma-like factor.

The DNA sequences encoding factors associated with virulence were found by the use of *in vivo* complementation assays, more particularly by complementation in a guinea pig model and in a mouse model. The *in vivo* genetic complementation systems utilized integrating shuttle cosmid libraries to identify potential virulence genes. Thus, the invention also provides techniques to identify a DNA sequence or sequences associated with virulence determinants in *M. tuberculosis* and *M. bovis* and similar DNA sequences in other tuberculosis complex strains and in strains of other mycobacterial species and in species of other pathogenic organisms.

Accordingly, embodiments of the invention include the following.

A method for identifying a DNA sequence or sequences associated with virulence determinants in *M. tuberculosis* and *M. bovis* and similar DNA sequences in other tuberculosis complex strains and in strains of other mycobacterial species and in species of other pathogenic organisms comprising the steps of:

a) preparing a genomic DNA library of the pathogenic organism;

b) constructing an integrating shuttle vector containing genomic inserts prepared in step a);

c) transforming via homologous recombination a population of avirulent organisms;

d) isolating the recombinants;

e) inoculating a subject with an adequate inoculum of the recombinants in order to select virulent recombinants;

f) isolating the virulent recombinants; and

g) identifying the DNA insert which confers virulence.

This method may be performed with individuals that are mice or guinea pigs.

An isolated polynucleotide comprised of a segment of less than 3kb that is essentially homologous to a mycobacterial DNA sequence associated with virulence in mycobacteria, wherein the mycobacterial DNA sequence encodes a sigma factor.

An isolated polynucleotide comprised of a segment of less than 3 kb that encodes a polypeptide or fragment thereof, wherein the polypeptide is associated with virulence in mycobacteria and is a sigma factor. The polypeptide may be essentially homologous to the polypeptide encoded in Figure 9.

An isolated polynucleotide comprised of at least 15 sequential nucleotides homologous to a sequence of polynucleotides in Figure 9.

A recombinant polynucleotide comprised of a sequence of at least 15 sequential nucleotides homologous to a sequence of polynucleotides in Figure 9.

A recombinant polynucleotide comprised of a segment of less than 3 kb that encodes a polypeptide or fragment thereof, wherein the polypeptide is associated with virulence in mycobacteria and is a sigma factor.

An expression vector comprised of the recombinant polynucleotide described above.

An isolated polynucleotide comprised of a linear segment of at least 15 nucleotides that is substantially homologous to mycobacterial DNA in a plasmid selected from the group consisting of pUHA1, pUHA2, pUHA3, pUHA4, pUHA5, pUHA6, pUHA7, pUHA8, pUHA9, pUHA11, pYUB352, pYUB353, and pYUB354.

A host cell comprised of any of the above-described isolated polynucleotides, including expression vectors.

A diagnostic kit comprised of a polynucleotide and a buffer packaged in suitable viale, wherein the polynucleotide is any of the above-described isolated polynucleotides.

An isolated polypeptide substantially homologous to a polypeptide associated with virulence in mycobacteria or a fragment thereof, wherein the mycobacterial polypeptide is a sigma factor. The mycobacterial polypeptide may be one that is encoded in a DNA sequence shown in Figure 9.

An isolated polynucleotide comprised of a segment of less than 3kb that is essentially homologous to a mycobacterial DNA sequence associated with avirulence in mycobacteria, wherein the mycobacterial DNA sequence encodes a sigma factor.

A method for producing an altered property in a wild-type bacterial strain other than *M. bovis* comprising mutagenizing a principal sigma factor in the bacteria, wherein the mutagenizing results in converting an arginine to a histidine in the principal sigma factor, and wherein the conversion occurs at a similar position to that present in *M. bovis* ATCC 35721. This method includes altering the virulence properties of the bacterial strain.

A method of using a bacterial strain prepared by the method described above, the method comprising preparing a vaccine by mixing a pharmacologically effective dose of the strain with a pharmaceutically acceptable suitable excipient.

Brief Description of the Drawings

Figure 1 is a schematic illustrating the strategy for recovering part of cosmid pUHA1 from *M. bovis* WAg300 which is a member of the *M. bovis* ATCC35721(pYUB178::*M. bovis* WAg200) library and which has increased virulence for guinea pigs. The diagrams are not to scale.

Figure 2 is a schematic showing the alignment of pUHA2-pUHA7 in linear form for comparison purposes beginning with the *Not*I site at position 2024 of pYUB178. Cosmids pUHA3-PUHA7 were isolated by colony hybridization using a probe of the 2 kb *Mlu*I fragment of PUHA2: M, *Mlu*I site; N, *Not*I site;

□, vector arm; _____, insert DNA from *M. bovis* Wag200.

Figure 3 is a restriction map of cosmid PUHA3 in linear form starting with the *Not*I site at position 2024 of pYUB178:h, *Nhe*I; M, *Mlu*I; N, *Not*I, X, *Xba*I.

Figures 4A-C represent a map of the integrating shuttle cosmid, pYUB178, and analysis of individual clones and pools of H37Ra(pYUB178::H37Rv).

Figure 4A shows the components that allow integration of pYUB178 into the mycobacterial genomes are *attP* and *int*. The pYUB178 cosmid contains an *E. coli ori*, the L5 attP, the L5 *int*, a kanamycin resistance gene, *aph*, derived from Tn*903*, lambda cos, and a unique cloning site, *Bcl*I.

Figure 4B is a schematic showing identification of the pYUB178/H37Rv junctional fragments within the chromosome of a H37Ra recombinant containing pYUB178::H37Rv DNA. *Pst*I-digested chromosomal DNA is separated by gel electrophoresis and hybridized with a labeled probe from pYUB178. The probe is the 1.1 kb *Dra*I/*Sap*I DNA fragment of pYUB178 that flanks the *Bcl*I cloning site. The integrated pYUB178::H37Rv cosmid can

be detected only by the presence of pYUB178-hybridizing DNA fragments. The *PstI* sites on either side of the H37Rv insert are fixed. Thus, the size of hybridizing DNA fragments varies with the H37Rv insert DNA.

Figure 4C are half-tones of gels showing individual H37Ra recombinants containing pYUB178::H37Rv cosmid clones were isolated from mouse lung tissue after spleen passage of recombinant pools, experiment J5P (see Table 9). Pools of H37Ra(pYUB178::H37Rv) were collected and passaged in broth culture. The chromosomal DNAs from pools and individual clones were isolated, digested with *PstI*, separated by agarose gel electrophoresis and transferred to a nylon filter to hybridize with the 1.1 kb DraI/SspI DNA fragment of pYUB178. Lanes 1-3, the H37Rv DNA junctional fragments of *in vivo*-selected individual clones of pool 2; lanes 4 and 5, the H37Rv DNA junctional fragments of members of pool 3, before (lane 4) and after (lane 5) *in vitro* passage.

Figures 5A-B shows the growth of *in vivo*-selected H37Ra(pYUB178::H37Rv) clones in mouse lung and spleen. Growth rates of clones $mc^2806$, H37Rv, and $mc^2816$ were measured and compared. The growth rate of $mc^2806$ is represented by solid squares on the solid lines, the growth rate of $mc^2816$ is represented by the open circles on the dotted lines, and the growth rate of H37Rv is represented by solid triangles on the dotted lines. These data are representative of three experiments. See text and Table 9, experiment J33, for experimental details.

Figure 5A shows growth in spleen.

Figure 5B shows growth in lung.

Figures 6A-B illustrate the retrieval of H37Rv-containing cosmids from the $mc^2806$ chromosome.

Figure 6A is a schematic illustrating the strategy used to retrieve the H37Rv insert DNA from the

integrated cosmids in H37Ra(pYUB178::H37Rv) recombinants.

Figure 6B is a half-tone of an autoradiograph showing a Southern hybridization of AseI and EcoRI digests of $mc^2806$ chromosomal DNA, or cosmid DNAs that were retrieved from the chromosome of $mc^2806$. The 436 bp AseI/BclI fragment of pYUB178 that contained cos was used as a probe. Lane 1, $mc^2806$ chromosomal DNA, lanes 2 to 17, DNA from sixteen individual retrieved cosmids.

Figure 7 is a graph showing the growth of H37Ra recombinants containing pYUB352-overlapping and -nonoverlapping cosmids. H37Ra was separately transformed with pYUB352-overlapping cosmids, pYUB353 and pYUB354, and with unrelated cosmids, pYUB355 and pYUB356. Growth of each recombinant was measured over a time course in mouse spleen. See Table 9, experiment J36. The growth of pYUB353- and pYUB354-containing H37Ra recombinants is represented by the small squares on the solid lines. The growth of $mc^2806$ is represented by the large squares on the solid lines. The growth of pYUB355- and pYUB356-containing H37Ra recombinants is represented by the small circles on the solid lines. The growth of $mc^2816$ is represented by the large circles on the dotted lines. The growth of H37Rv is represented by the triangles on the dotted lines.

Figures 8A-C represent the restriction map of the *ivg* region of H37Rv DNA in pYUB352-overlapping cosmids. Restriction digests of pYUB352, pYUB353, and pYUB354 were performed with EcoRI and HindIII.

Figure 8A is a half-tone reproduction of gels showing digested DNA fragments which were separated by agarose gel electrophoresis.

Figure 8B is a half-tone reproduction of gels showing DNA fragments which were hybridized to the AseI fragment of pYUB352 that included its entire H37Rv insert with flanking pYUB178 DNA sequences. The arrows point to

DNA fragments that hybridize to pYUB178 DNA probes. These bands are junctional fragments. Lanes 1-3 are digests of pYUB352, lanes 4-6 are digests of pYUB353, and lanes 7-9 are digests of pYUB354. Lanes 1, 4, and 7 show *Eco*RI digestion patterns, lanes 2, 5, and 8 show *Eco*RI and *Hind*III double digestion patterns, and lanes 3, 6, and 9 show *Hind*III digestion patterns.

Figure 8C is a schematic illustrating data gathered from these molecular analyses and the functional analyses shown in Figure 7 allowed the construction of the physical map of the *ivg* region of H37Rv that is present in cosmids pYUB352, pYUB353, and pYUB354. A=*Ase*I, E=*Eco*RI, H=*Hind*III.

Figure 9 and 9a is comprised of four sheets. Figure 9 shows the nucleotide sequence of the coding strand of the 2745 bp fragment that restores virulence to *M. bovis* ATCC35721. Figure 9a shows the same as in Figure 9 together with a 530 amino acid sequence translated from the largest ORF.

Figure 10A is comprised of two sheets showing the results of a PileUp comparison of known principal sigma factors from *Streptomyces coelicolor* (GenBank Accession Nos. χ52980, χ52981, χ52983) and *Streptomyces griseus* (GenBank Accession No. L08071) with the translation of the largest ORF of the 2000 bp contig from the *M. bovis* virulence restoring factor, rpoV, that restores virulence to *M. bovis* ATCC35721.

Figure 11 presents the results of a GAP comparison of *Streptomyces griseus* principal sigma factor (Peptide translation of GenBank accession No. L08071 from nucleotide numbers 570 to 1907, which is the coding sequence of the *hrdB* gene) with peptide translation of the large ORF of the approximately 3 kb DNA fragment from *M. bovis* associated with virulence.

Figure 12a-1 and 12a-2 (SEQ ID NO:13 and SEQ ID NO:14) is comprised of two sheets showing the large ORF of the *M. bovis* WAg200 sequence which begins with GTG at position 835-837.

Figure 12 (SEQ ID NO:8 through SEQ ID NO:12) presents a comparison of putative principal sigma factors of three *M. tuberculosis* complex strains and two *Streptomyces sp.*

Detailed Description of the Invention

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Sambrook, Fritsch, and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, Second Edition (1989), OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait Ed., 1984), the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J.M. Miller and M.P. Calos eds. 1987), HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, (D.M. Weir and C.C. Blackwell, Eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Siedman, J.A. Smith, and K. Struhl, eds., 1987), and CURRENT PROTOCOLS IN IMMUNOLOGY (J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach and W. Strober, eds., 1991). All patents, patent applications, and publications mentioned herein, both *supra* and *infra*, are incorporated herein by reference.

The present invention provides polynucleotides that are associated with virulence in members of the genus mycobacteria, and particularly in members of the mycobacterial complex. Virulence is the relative capacity of a pathogen to overcome body defenses; it is

also the relative ability to cause disease in an infected host. In gram-negative bacterial pathogens, virulence is

generally determined by a multiplicity of traits that endow the pathogen with its ability to exploit anatomical weaknesses and overcome the immune defenses of the host. It is expected that a similar multiplicity of traits determines the virulence of pathogenic mycobacteria. Properties associated with virulence in microorganisms include those listed in Table 1.

Table 1. Properties associated with virulence

1. Infectious; capable of being spread from one individual to another.
2. Capable of entering mammalian host cells.
3. Capable of surviving or escaping phagocyte cellular defenses.
4. Capable of multiplying in host cells.
5. Capable of spreading from one infected cell to an uninfected cell.
6. Capable of causing cell injury that results in pathology.

In addition, a virulent organism may be capable of killing the infected host.

By mycobacteria is meant the genus that includes the species *M. phlei, M. smegmatis, M. africanum, M. fortuitum, M. marinum, M. ulcerans, M. tuberculosis, M. bovis, M. microti, M. avium, M. paratuberculosis, M. leprae, M. lepraemurium, M. intracellulare, M. scrofulaceum, M. xenopi, M. genavense, M. kansasii, M. simiae, M. szulgai, M. haemophilum, M. asiaticum, M. malmoense, and M. shimoidei*. Of particular interest are the members of the tuberculosis complex, including *M. tuberculosis, M. bovis, M. africanum and M. microti*.

As used herein, the term "virulence factor encoding sequence" denotes a polynucleotide sequence that encodes a product that is associated with virulence in a

member of the mycobacterial species. This term is encompassed within the term a "sequence associated with virulence" that denotes that a polynucleotide sequence that confers a trait associated with virulence on an avirulent mycobacterium, whether or not the polynucleotide encodes a product. In particular, the virulence associated sequences of the present invention are those that confer one or more traits associated with virulence and have a high degree of homology, i.e., at least about 70% overall homology, preferably at least about 80% overall homology, even more preferably at least about 90% overall homology, to the mycobacterial polynucleotides described herein. Methods of determining homology between sequences are known in the art, and include, for example, direct comparison of sequences, and hybridization assays.

The sequence of one of the mycobacterial DNAs associated with virulence, isolated from *M. bovis*, is shown in Figure 9. This DNA contains several contigs and an open reading frame (ORF) that based upon amino acid sequence homology in certain regions, encodes a polypeptide that is a putative sigma factor. Portions or all of fragment of which the ORF is part is in plasmids pUHA1, pUHA2, pUHA3, pUHA4, pUHA5, pUHA6, PUHA7, pUHA8, pUHA9, or pUHA11. A particular embodiment of the invention is an isolated or recombinant polynucleotide that is comprised of all or segment of the ORF encoding the sigma factor.

Virulence is also associated with the mycobacterial sequences present in pYUB352, pYUB353, and pYUB354. Thus, the isolated and recombinant polynucleotides may also be comprised of sequences homologous to the mycobacterial DNA in these plasmids.

The DNA sequences upon which the polynucleotides of the invention are based were obtained

by the use of *in vivo* virulence complementation assays. A method for identifying virulence determinants by genetic complementation *in vivo* was discovered that requires: (i) two strains that are genetically similar; (ii) a phenotype associated with virulence; and (iii) gene transfer systems.

Cosmid genomic libraries of virulent mycobacterial strains of *M. tuberculosis* and *M. bovis* were constructed in an integrating cosmid vector. An example of an integrating cosmid vector is pYUB178, described by Lee et al. (1991), Proc. Natl. Acad. Sci. USA, 88:3111-3115 and Pascopella et al. (1994), Infect. Immun. 62:1313-1319. The integrating vector, approximately 5 kb long, can accommodate 40-45 kb of DNA and uses the site-specific integration system of mycobacteriophage L5 to integrate recombinant DNA into a unique *attB* site of the mycobacterial chromosome. This vector thus can represent more than 95% of the entire mycobacterial genome in as few as about 300 clones. The recombinant DNA introduced in single copy is stably maintained in mycobacterial cells in the absence of antibiotic selection, even when the strain is passed through animals. Thus, use of this vector reduced the number of clones that needed to be screened, and ensured that cloned genes were not lost during animal passage.

The genomic libraries in the integrating cosmid vector were introduced into corresponding avirulent strains of mycobacteria. Methods of introducing polynucleotides into cells are known in the art, and include, for example, electroporation, transduction and transformation. In order to select for virulent mycobacteria the resulting libraries of recombinant clones were injected into animals, i.e., mice or guinea pigs. It is thought that clones that restore virulence may have a selective advantage and thus be enriched for

in the injected animals. In the mouse complementation assay, avirulent mutants cause a self-limiting infection while virulent mycobacterial strains multiply more rapidly, and in high challenge doses cause death. Similarly, in the guinea pig complementation assay, avirulent mutants cause a self-limiting infection. However, virulence in guinea pigs can be assessed by the sites in which gross lesions are found. When avirulent strains of mycobacteria are inoculated subcutaneously in a flank, these strains are not sufficiently virulent to pass through the lymph nodes draining the injection site and enter the systemic circulation in sufficient numbers to cause gross lesions to occur in the spleen. This is contrasted to virulent strains, which under the same inoculation conditions do give rise to spleen (and lung) lesions. Examples of assay systems for comparing avirulent and corresponding virulent mutants of *M. tuberculosis* and *M. bovis* are described in the Examples.

Clones of mycobacteria that had been rendered virulent by the integration of a polynucleotide encoding a virulence factor were isolated. Portions of the integrated virulence determining cosmid were isolated from the clones by restriction enzyme digestion, and the fragments were reinserted into the integrating vector and assayed for virulence factor activity using *in vivo* complementation assays. These assays led to the identification of mycobacterial DNA encoding polypeptides associated with virulence. In the case of *M. bovis*, the sequence of a fragment of mycobacterial DNA of approximately 3 kb in a clone designated pUHA11 was determined. A comparison of GenBank sequences with the amino acids encoded in the fragment, and particularly within a large ORF and an adjacent contig, showed a significant degree of homology with sigma factors from other microorganisms, indicating that the large ORF

encodes a putative sigma factor. On the basis of this homology and the ability of the WAg200 gene to confer a virulence phenotype we have named the gene, *rpoV*. The high degree of homology between the principal sigma factors of *Streptomyces* sp. and the putative sigma factors from the *M. tuberculosis* complex may reflect their evolutionary relationship and the fact that both these genera have DNA with a high guanine plus cytosine percentage.

A comparison of the homologous DNA sequences from *M. bovis* WAg200 and the DNA sequence from the attenuated *M. bovis* ATCC35721 indicated that the latter had no sequence differences upstream of the ORF but had two point differences in the coding sequence. One of these differences was also present in the virulent strain *M. tuberculosis* Erdman but the other difference, which caused an arginine to histidine change at position 522, was not found in any of the virulent strains analyzed. Thus we deduce that this is the likely mutation that causes *M. bovis* ATCC35721 to become avirulent. This position is highly conserved among principal sigma factors and their homologues and the region in which it occurs has the characteristics of a helix-turn-helix motif and is believed to be involved in -35 sequence recognition. See Lonetto, M., Gribskov, M. and Gross, C.A., (1992) J. Bact. 174: 3843-3849. Thus, as used herein, the term "similar position to that present in *M. bovis* ATCC35721" in reference to arginine to histidine conversion in a bacterial strain with a mutagenized principal sigma factor contemplates one in a region that is highly conserved among principal sigma factors and their homologues and one that has the characteristics of a helix-turn-helix motif and is believed to be involved in -35 sequence recognition.

While the virulence assays initially were used to isolate the polynucleotides described herein, they may also be used to determine whether polynucleotides constructed from the information and sequences provided herein and factors transcribed and/or translated therefrom are associated with virulence in mycobacteria, and particularly in *M. bovis* or *M. tuberculosis*.

One embodiment of the invention is an isolated polynucleotide comprised of a sequence associated with virulence in mycobacteria. Another embodiment of the invention is an isolated polynucleotide comprised of a sequence associated with avirulence in mycobacteria. As used herein the term "polynucleotide" refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA and RNA. It also includes known types of modifications, for example, labels which are known in the art (*e.g.*, Sambrook, *et al.*), methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.), those containing pendant moieties, such as, for example, proteins (including for *e.g.*, nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (*e.g.*, acridine, psoralen, etc.), those containing chelators (*e.g.*, metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide. Polynucleotides include both sense and antisense strands. Recombinant nucleic acids comprising sequences

otherwise not naturally occurring with the designated mycobacterial sequence are also provided by this invention. Although the wild type sequence may be employed, the wild type sequence will often be altered, e.g., by deletion, substitution, or insertion.

The nucleic acid sequences used in this invention will usually comprise at least about 5 codons (15 nucleotides), more usually at least about 7 to 15 codons, and most preferably at least about 35 codons. One or more introns may also be present. This number of nucleotides is usually about the minimal length required for a successful probe that would hybridize specifically with such a sequence.

Techniques for nucleic acid manipulation are described generally, for example, in Sambrook et al., ibid., or Ausubel et al., ibid.. Reagents useful in applying such techniques, such as restriction enzymes and the like, are widely known in the art and commercially available from such vendors as New England BioLabs, Boehringer Mannheim, Amersham, Promega Biotec, U. S. Biochemicals, New England Nuclear, and a number of other sources.

The polynucleotides of the invention will have substantial homology or similarity to the DNAs disclosed herein that are associated with virulence or with avirulence in mycobacteria. A nucleic acid or fragment thereof is "substantially homologous" (or "substantially similar") to another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95 to 98% of the nucleotide bases.

Alternatively, a nucleic acid or fragment (or its complementary strand) is substantially homologous (or similar) with a DNA associated with virulence or with avirulence in mycobacteria when they are capable of hybridizing under selective hybridization conditions. Selectivity of hybridization exists when hybridization occurs which is substantially more selective than total lack of specificity. Typically, selective hybridization will occur when there is at least about 65% homology over a stretch of at least about 14 nucleotides, preferably at least about 70%, more preferably at least about 75%, and most preferably at least about 90%. See, Kanehisa (1984) Nuc. Acids Res. 12:203-213. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least about 17 nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration (e.g., NaCl), temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30° C, typically in excess of 37°, and preferably in excess of 45°. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur and Davidson (1968) J. Mol. Biol. 31:349-370.

The polynucleotides of the invention are isolated or substantially purified. An "isolated" or "substantially pure" or "purified" nucleic acid is a nucleic acid, e.g., an RNA, DNA, or a mixed polymer, which is substantially separated from other mycobacterial components that naturally accompany the sequences associated with virulence, e.g., ribosomes, polymerases, and many other mycobacterial polynucleotides such as RNA and other chromosomal sequences. The term embraces a nucleic acid sequence which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogues or analogues biologically synthesized by heterologous systems.

The term "recombinant polynucleotide" as used herein intends a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature; or (2) is linked to a polynucleotide other than that to which it is linked in nature; and (3) does not occur in nature. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions.

In some embodiments of the invention the polynucleotides encode a polypeptide associated with virulence or with avirulence. A nucleic acid is said to

"encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the polypeptide or a fragment thereof. The anti-sense strand of such a nucleic acid is also said to encode the sequence.

Also contemplated within the invention are expression vectors comprised of a sequence encoding a polypeptide associated with virulence. Expression vectors generally are replicable polynucleotide constructs that encode a polypeptide operably linked to suitable transcriptional and translational regulatory elements. Examples of regulatory elements usually included in expression vectors are promoters, enhancers, ribosomal binding sites, and transcription and translation initiation and termination sequences. These regulatory elements are operably linked to the sequence to be translated. A nucleic acid sequence is operably linked when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression. Generally, operably linked means that the DNA sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. The regulatory elements employed in the expression vectors containing a polynucleotide encoding a virulence factor are functional in the host cell used for expression.

The polynucleotides of the present invention may be prepared by any means known in the art. For example, large amounts of the polynucleotides may be produced by replication in a suitable host cell. The natural or synthetic DNA fragments coding for a desired fragment will be incorporated into recombinant nucleic

acid constructs, typically DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the DNA constructs will be suitable for autonomous replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to and integration within the genome of a cultured insect, mammalian, plant or other eukaryotic cell lines. The purification of nucleic acids produced by the methods of the present invention are described, e.g., in Sambrook et al. (1989) or Ausubel et al. (1987 and periodic updates).

The polynucleotides of the present invention may also be produced by chemical synthesis, e.g., by the phosphoramidite method described by Beaucage and Carruthers (1981) Tetra. Letts. 22:1859-1862 or the triester method according to Matteucci et al. (1981) J. Am. Chem. Soc. 103:3185, and may be performed on commercial automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strand together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

DNA constructs prepared for introduction into a prokaryotic or eukaryotic host will typically comprise a replication system recognized by the host, including the intended DNA fragment encoding the desired polypeptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding segment. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as

ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences. Secretion signals from polypeptides secreted from the host cell of choice may also be included where appropriate, thus allowing the protein to cross and/or lodge in cell membranes, and thus attain its functional topology or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art and discussed, for example, in Sambrook et al. (1989) or Ausubel et al. (1987).

The selection of an appropriate promoter and other necessary vector sequences will be selected so as to be functional in the host, and may, when appropriate, include those naturally associated with mycobacterial genes. Examples of workable combinations of cell lines and expression vectors are described in Sambrook et al., 1989 or Ausubel et al., 1987); see also, e.g., Metzger et al. 1988), Nature 334:31-36. Many useful vectors are known in the art and may be obtained from such vendors as Stratagene, New England Biolabs, Promega Biotech, and others. Promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters may be used in prokaryotic hosts. Useful yeast promoters include the promoter regions for metallothionein, 3-phosphoglycerate kinase or other glycolytic enzymes such as enolase or glyceraldehyde-3-phosphate dehydrogenase, enzymes responsible for maltose and galactose utilization, and others. Suitable vectors and promoters for use in yeast expression are further described in Hitzeman et al. EP 73,657A. Appropriate nonnative mammalian promoters might include the early and late promoters from SV40 (Piers et al. (1978) Nature 273:113) or promoters derived from murine moloney leukemia virus, mouse mammary tumor virus, avian sarcoma

viruses, adenovirus II, bovine papilloma virus or polyoma. In addition, the construct may be joined to an amplifiable gene (e.g., DHFR) so that multiple copies of the gene may be made. For appropriate enhancer and other expression control sequences see also Enhancers and Eukaryotic Gene Expression, Cold Spring Harbor Press, N.Y. (1983).

While such expression vectors may replicate autonomously, they may less preferably replicate by being inserted into the genome of the host cell, by methods well known in the art.

Expression and cloning vectors will likely contain a selectable marker, a gene encoding a protein necessary for the survival or growth of a host cell transformed with the vector. The presence of this gene ensures the growth of only those host cells which express the inserts. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxic substances, e.g. ampicillin, neomycin, methotrexate, etc.; (b) complement auxotrophic deficiencies; or (c) supply critical nutrients not available from complex media, e.g. the gene encoding D-alanine racemase for Bacilli. The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art.

The vectors containing the nucleic acids of interest can be transcribed *in vitro* and the resulting RNA introduced into the host cell by well known methods (e.g., by injection. See, T. Kubo et al., FEBS Lett. 241:119 (1988)), or the vectors can be introduced directly into host cells by methods well known in the art, which vary depending on the type of cellular host, including electroporation; transfection employing calcium chloride, rubidium chloride calcium phosphate, DEAE-dextran, or other substances; microprojectile

bombardment; lipofection; infection (where the vector is an infectious agent, such as a retroviral genome); and other methods. See generally, Sambrook et al. (1989) and Ausubel et al. (1987). The cells into which have been introduced nucleic acids described above are meant to also include the progeny of such cells.

Large quantities of the nucleic acids and polypeptides of the present invention may be prepared by expressing the nucleic acids or portions thereof in vectors or other expression vehicles in compatible prokaryotic or eukaryotic host cells. The most commonly used prokaryotic hosts are strains of Escherichia coli, although other prokaryotes, such as Bacillus subtilis or Pseudomonas may also be used.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, amphibian or avian species, may also be useful for production of the proteins of the present invention. Propagation of mammalian cells in culture is per se well known. See, Tissue Culture, Kruse and Patterson, ed., Academic Press (1973). Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and WI38, BHK, and COS cell lines, although it will be appreciated by the skilled practitioner that other cell lines may be appropriate, e.g., to provide higher expression, desirable glycosylation patterns, or other features.

Clones are selected by using markers depending on the mode of the vector construction. The marker may be on the same or a different DNA molecule, preferably the same DNA molecule. The transformant may be screened or, preferably, selected by any of the means well known in the art, e.g., by resistance to such antibiotics as ampicillin, tetracycline.

Also included within the invention are isolated or recombinant polynucleotides that bind to the regions of the mycobacterial chromosome containing sequences that are associated with virulence, including antisense and triplex forming polynucleotides. As used herein, the term "binding" refers to an interaction or complexation between an oligonucleotide and a target nucleotide sequence, mediated through hydrogen bonding or other molecular forces. The term "binding" more specifically refers to two types of internucleotide binding mediated through base-base hydrogen bonding. The first type of binding is "Watson-Crick-type" binding interactions in which adenine-thymine (or adenine-uracil) and guanine-cytosine base-pairs are formed through hydrogen bonding between the bases. An example of this type of binding is the binding traditionally associated with the DNA double helix and in RNA-DNA hybrids; this type of binding is normally detected by hybridization procedures.

The second type of binding is "triplex binding". In general, triplex binding refers to any type of base-base hydrogen bonding of a third polynucleotide strand with a duplex DNA (or DNA-RNA hybrid) that is already paired in a Watson-Crick manner.

The invention also includes recombinant host cells comprised of any of the above described polynucleotides that contain a sequence associated with virulence in mycobacteria, including those encoding a polypeptide, particularly a polypeptide that is substantially homologous to the polypeptide encoded in Figure 9, or a fragment thereof, or an analog thereof.

The polynucleotides of the invention may be inserted into the host cell by any means known in the art, including for example, transformation, transduction, and electroporation. As used herein, "recombinant host

cells", "host cells", "cells", "cell lines", "cell cultures", and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refer to cells which can be, or have been, used as recipients for recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transformed. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

"Transformation", as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transduction, f-mating or electroporation. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host cell genome.

The polynucleotides of the invention that are essentially homologous to sequences associated with virulence, shown in Figure 9, and in plasmids pUHA1, pUHA2, pUHA3, pUHA4, pUHA5, pUHA6, pUHA7, pUHA11 and pUHA16, and in plasmids pYUB352, pYUB353, pYUB354 are of use in the detection of virulent forms of mycobacteria in biological samples. As used herein, a "biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, blood cells, tumors, organs, and also samples of <u>in vitro</u> cell culture constituents (including but not limited to conditioned medium resulting from the growth of cells in

cell culture medium, putatively virally infected cells, recombinant cells, and cell components).

Using the disclosed portions of the isolated polynucleotides associated with virulence as a basis, oligomers of approximately 8 nucleotides or more can be prepared, either by excision from recombinant polynucleotides or synthetically, which hybridize with the mycobacterial sequences in the plasmids and are useful in identification of mycobacteria with the virulence associated trait. The probes for polynucleotides associated with virulence are a length which allows the detection of the virulence associated sequences by hybridization. While 6-8 nucleotides may be a workable length, sequences of 10-12 nucleotides are preferred, and at least about 20 nucleotides appears optimal. These probes can be prepared using routine methods, including automated oligonucleotide synthetic methods. For use as probes, complete complementarity is desirable, though it may be unnecessary as the length of the fragment is increased.

For use of such probes as diagnostics, the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids contained therein. The resulting nucleic acid from the sample may be subjected to gel electrophoresis or other size separation techniques; alternatively, the nucleic acid sample may be dot blotted without size separation. The probes are usually labeled. Suitable labels, and methods for labeling probes are known in the art, and include, for example, radioactive labels incorporated by nick translation or kinasing, biotin, fluorescent probes, and chemiluminescent probes. The nucleic acids extracted from the sample are then treated with the labeled probe under hybridization conditions of suitable stringencies.

The probes can be made completely complementary

to the virulence encoding polynucleotide. Therefore, usually high stringency conditions are desirable in order to prevent false positives. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, length of time, and concentration of formamide. These factors are outlined in, for example, Maniatis, T. (1982).

It may be desirable to use amplification techniques in hybridization assays. Such techniques are known in the art and include, for example, the polymerase chain reaction (PCR) technique described which is by Saiki et al. (1986), by Mullis, U.S. Patent No. 4,683,195, and by Mullis et al. U.S. Patent No. 4,683,202.

The probes can be packaged into diagnostic kits. Diagnostic kits include the probe DNA, which may be labeled; alternatively, the probe DNA may be unlabeled and the ingredients for labeling may be included in the kit in separate containers. The kit may also contain other suitably packaged reagents and materials needed for the particular hybridization protocol, for example, standards, as well as instructions for conducting the test.

Polypeptides encoded within the sequences associated with virulence, and fragments and analogs thereof are also included as embodiments of the invention. The polypeptide encoded in the large ORF in Figure 9 is a putative sigma factor; thus, the intact polypeptide may exhibit the following biological activities: (1) binding to mycobacterial core RNA polymerase, (b) activation of promoter recognition;, and may include (c) DNA melting and (d) inhibition of nonspecific transcription. Methods to determine these biological functions are known in the art, and for

example are reviewed in J.D. Helmann and M.J. Chamberlin, Ann. Rev. Biochem. (1988) 57, 839-872. Also included as a biological activity of any specific polypeptide is the binding of the polypeptide to an antibody that is directed to one or more epitopes on that polypeptide. The invention includes polypeptides and analogs or fragments thereof that are essentially homologous to the polypeptide encoded in the large ORF in Figure 9, and exhibit at least one of the biological activities associated with sigma factor, or alternatively, inhibits at least one of the biological activities associated with sigma factor.

The term "polypeptide" refers to a polymer of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as the modifications known in the art, both naturally occurring and non-naturally occurring.

Ordinarily, the polypeptides of the present invention will be at least about 50% homologous to the polypeptide encoded in the large ORF of Figure 9, designated herein as "virulence associated sigma factor 1" (also referred to herein as "*rpoV*"), preferably in excess of about 90%, and, more preferably, at least about 95% homologous. Also included are proteins encoded by DNA which hybridize under high or low stringency conditions, to nucleic acids encoding virulence associated sigma factor 1, as well as closely related

polypeptides or proteins retrieved by antisera to virulence associated sigma factor 1.

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acids, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues.

The term "substantial homology" or "substantial identity", when referring to polypeptides, indicates that the polypeptide or protein in question exhibits at least about 30% identity with an entire naturally occurring protein or a portion thereof, usually at least about 70% identity, and preferably at least about 95% identity.

Homology, for polypeptides, is typically measured using sequence analysis software. See, e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wisconsin 53705. Protein analysis software matches similar sequences using measure of homology assigned to various substitutions, deletions, substitutions, and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, often at least about 7 amino acids, typically at least about 9 to 13 amino acids, and, in various embodiments, at least about 17 or more amino acids.

The terms "isolated," "substantially pure," and "substantially homogenous" are used interchangeably to describe a protein or polypeptide which has been

separated from components which naturally accompany it. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and preferably will be over about 99% pure. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

A protein is considered to be isolated when it is separated from the contaminants which accompany it in its natural state. Thus, a polypeptide which is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components.

The present invention provides polypeptides which may be purified from mycobacteria as well as from other types of cells transformed with recombinant nucleic acids encoding these proteins. Such protein purification can be accomplished by various methods well known in the art, and include those described, e.g., in Guide to Protein Purification, ed. M. Deutscher, vol. 182 of Methods in Enzymology (Academic Press, Inc.: San Diego, 1990) and R. Scopes, Protein Purification: Principles and Practice, Springer-Verlag: New York, 1982.

If necessary, the amino acid sequence of the proteins of the present invention can be determined by protein sequencing methods well known in the art.

The present invention also provides for polypeptides or fragments thereof which are substantially

homologous to the primary structural sequence of the virulence associated sigma factor 1 (also called *rpoV*). The present invention also embraces *in vivo* or *in vitro* chemical and biochemical modifications that incorporate unusual amino acids. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labelling, e.g., with radionuclides, various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labelling polypeptides and of substituents or labels useful for such purposes are well known in the art and include radioactive isotopes such as $^{32}P$, ligands, which bind to labeled antiligands (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods of labelling polypeptides are well known in the art. See, e.g., Molecular Cloning: A Laboratory Manual, 2nd ed., Vol. 1-3, ed. Sambrook, et al., Cold Spring Harbor Laboratory Press (1989) or Current Protocols in Molecular Biology, ed. F. Ausubel et al., Greene Publishing and Wiley-Interscience: New York (1987 and periodic updates).

Besides substantially full-length polypeptides, the present invention provides for fragments of the polypeptides capable of binding to antibodies directed to virulence associated sigma factor 1. As used herein, the term fragment or segment, as applied to a polypeptide, will ordinarily be at least about 5 to 7 contiguous amino acids, typically at least about 9 to 13 contiguous amino acids, and most preferably at least about 20 to 30 or more contiguous amino acids.

The present invention also provides for fusion polypeptides comprising the virulence associated sigma factor 1 or fragments thereof. Homologous polypeptides may be fusions between two or more sequences derived from the virulence associated sigma factor 1 or between the sequences of the virulence associated protein and a related protein. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. See, e.g., Godowski et al. (1988) Science 241:812-816.

Fusion proteins will typically be made by recombinant nucleic acid methods, but may be chemically synthesized. Techniques for synthesis of polypeptides are described, for example, in Merrifield (1963) J. Amer. Chem. Soc. 85:2149-2156.

The polypeptides of the present invention may be used in the preparation of vaccines to treat and/or prevent diseases associated with mycobacterial infections. "Treatment" as used herein refers to prophylaxis and/or therapy.

The polypeptides can be prepared as discrete entities or incorporated into a larger polypeptide, and may find use as described herein. The immunogenicity of the epitopes of the polypeptides of the invention may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B surface antigen. See, e.g., U.S. Pat. No. 4,722,840. Vaccines may be prepared from one or more immunogenic polypeptides derived from virulence associated polypeptides, and more particularly from virulence associated sigma factor 1.

The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredients, is known to one skilled in the art. Typically, such

vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing an rpoV antigenic sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of

administration include suppositories and, in some cases, oral formulations or formulations suitable for distribution as aerosols. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or with organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5 micrograms to 250 micrograms of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection

desired. Precise amounts of active ingredient required to be administered may depend on the judgment of the practitioner and may be peculiar to each subject.

The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgment of the practitioner.

In addition, the vaccine containing the immunogenic mycobacterial antigen(s) may be administered in conjunction with other immunoregulatory agents, for example, immune globulins, as well as antibiotics.

The immunogenic virulence associated antigens may be used for the preparation of antibodies. The immunogenic polypeptides prepared as described above are used to produce antibodies, including polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized with an immunogenic polypeptide bearing an *rpoV* epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an *rpoV* epitope contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987).

Monoclonal antibodies directed against *rpoV* epitopes can also be readily produced by one skilled in

the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al. (1980); Hammerling et al. (1981); Kennett et al. (1980); see also, U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against *rpoV* epitopes can be screened for various properties; i.e., for isotype, epitope affinity, etc.

Antibodies, both monoclonal and polyclonal, which are directed against *rpoV* epitopes are particularly useful in diagnosis, and those which are neutralizing may be useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies. Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is desired. See, for example, Nisonoff, A., et al. (1981) and Dreesman et al. (1985). Techniques for raising anti-idiotype antibodies are known in the art. See, for example, Grzych (1985), MacNamara et al. (1984), and Uytdehaag et al. (1985). These anti-idiotype antibodies may also be useful for treatment, vaccination and/or diagnosis of mycobacterial infections, as well as for an elucidation of the immunogenic regions of *rpoV* antigens.

Both the virulence associated polypeptides and antibodies to them are useful in immunoassays to detect presence of antibodies to mycobacteria, or the presence of the virulence associated antigens, and particularly the presence of virulence associated *rpoV* in biological

samples. Design of the immunoassays is subject to a great deal of variation, and many formats are known in the art. The immunoassay will utilize at least one epitope derived from a virulence associated polypeptide, and particularly virulence associated *rpoV*. In one embodiment, the immunoassay uses a combination of epitopes derived from the virulence associated polypeptide. These epitopes may be derived from the same or from different bacterial polypeptides, and may be in separate recombinant or natural polypeptides, or together in the same recombinant polypeptides. An immunoassay may use, for example, a monoclonal antibody directed towards a virulence associated polypeptide epitope(s), a combination of monoclonal antibodies directed towards epitopes of one mycobacterial antigen, monoclonal antibodies directed towards epitopes of different mycobacterial antigens, polyclonal antibodies directed towards the same antigen, or polyclonal antibodies directed towards different antigens. Protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, enzymatic, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Typically, an immunoassay for an antibody(s) to a virulence associated polypeptide, and particularly to virulence associated *rpoV* will involve selecting and preparing the test sample suspected of containing the antibodies, such as a biological sample, then incubating it with an antigenic (i.e., epitope-containing) virulence

associated polypeptide(s) under conditions that allow antigen-antibody complexes to form, and then detecting the formation of such complexes. Suitable incubation conditions are well known in the art. The immunoassay may be, without limitations, in a heterogenous or in a homogeneous format, and of a standard or competitive type.

In a heterogeneous format, the polypeptide is typically bound to a solid support to facilitate separation of the sample from the polypeptide after incubation. Examples of solid supports that can be used are nitrocellulose (e.g., in membrane or microtiter well form), polyvinyl chloride (e.g., in sheets or microtiter wells), polystyrene latex (e.g., in beads or microtiter plates, polyvinylidine fluoride (known as Immulon), diazotized paper, nylon membranes, activated beads, and Protein A beads. For example, Dynatech Immulon [1] or Immulon [2] microtiter plates or 0.25 inch polystyrene beads (Precision Plastic Ball) can be used in the heterogeneous format. The solid support containing the antigenic polypeptide is typically washed after separating it from the test sample, and prior to detection of bound antibodies. Both standard and competitive formats are known in the art.

Complexes formed comprising anti-rpoV antibody (or, in the case of competitive assays, the amount of competing antibody) are detected by any of a number of known techniques, depending on the format. For example, unlabeled anti-virulence associated polypeptide antibodies in the complex may be detected using a conjugate of antixenogeneic Ig complexed with a label, (e.g., an enzyme label).

In immunoassays where the virulence associated polypeptides are the analyte, the test sample, typically a biological sample, is incubated with antibodies

directed against the virulence associated polypeptide under conditions that allow the formation of antigen-antibody complexes. It may be desirable to treat the biological sample to release putative bacterial components prior to testing. Various formats can be employed. For example, a "sandwich assay" may be employed, where antibody bound to a solid support is incubated with the test sample; washed; incubated with a second, labeled antibody to the analyte, and the support is washed again. Analyte is detected by determining if the second antibody is bound to the support. In a competitive format, which can be either heterogeneous or homogeneous, a test sample is usually incubated with antibody and a labeled, competing antigen is also incubated, either sequentially or simultaneously. These and other formats are well known in the art.

Also included as an embodiment of the invention is an immunoassay kit comprised of one or more polypeptides of the invention, or antibodies to a polypeptide associated with virulence, and a buffer, packaged in suitable containers.

In addition, compounds which block the activity of virulence factor associated polypeptides and particularly virulence associated rpoV, may be prepared utilizing the sequence information of provided herein. This is performed by overexpressing the polypeptide, purifying the polypeptide, and then performing X-ray crystallography on the purified virulence associated polypeptide to obtain its molecular structure. Next, compounds are created which have similar molecular structures to all or portions of the polypeptide or its substrate. The compounds are then combined with the polypeptide and attached thereto so as to block one or more of its biological activities.

The polynucleotides of the invention may also be used to produce or improve live attenuated or killed tuberculosis vaccines. For example a vaccine strain may be produced by mutating a virulence associated polynucleotide, and particularly one encoding virulence associated sigma factor 1. The mutated strain may then be formulated into a vaccine and administered to treat mycobacterial infections. In addition, virulence associated polynucleotides may be added to BCG vaccine strains to provide attenuated mutant tuberculosis vaccines.

The invention also encompasses a new approach for determining factors associated with virulence or other properties of interest in other genera of bacteria by showing that an arginine to histidine change near the C-terminal end of a principal sigma factor, and in particular at the equivalent site to that which occurs in *M. bovis* AtCC35721, is not lethal but causes an alteration in the specificity of promotion of the sigma factor. Such a change could be engineered in the principal sigma factor in species of other genera of bacteria using techniques known in the art, including for example, site directed mutagenesis and homologous recombination. Identification and subsequent investigation of the genes whose promotion is altered by such a change could be performed using techniques known to one of skill in the art, for example, comparative protein electrophoresis, partial protein sequencing and reverse genetic methods. One might also use, for example, *in vivo* methods for identifying the level of promotion of different promoters in the presence of normal and altered sigma factors. The results of these studies should reveal genes whose promotion changes significantly when promoted by an altered principal sigma factor. Such genes may be potential targets for new

drugs or they could be targets for inactivation to generate new strains for use in vaccines or strains with other desirable properties.

The following examples are provided only for illustrative purposes, and not to limit the scope of the present invention. In light of the present disclosure numerous embodiments within the scope of the claims will be apparent to those of ordinary skill in the art.

Example 1

ISOLATION OF A VIRULENCE FACTOR OF MYCOBACTERIA USING A GUINEA PIG COMPLEMENTATION ASSAY.

Virulent tuberculosis complex strains were cultured as described previously (Collins and de Lisle 1984). Mycobacterial species were identified by standard methods.

For preparation of genomic DNA, tuberculosis complex strains were grown on standard mycobacterial media, harvested into buffer and inactivated by heating.

Genomic DNA was prepared form the organisms and partially digested with a range of concentrations of *Sau3AI*. Fragments of 30-50 kb from these digestions were prepared using sucrose gradient centrifugation and ligated to *BclI*-digested pYUB178 DNA that had been treated with calf intestinal phosphatase. The ligation mixture was *in vitro*-packaged into γ phage heads and transduced into *Escherichia coli*. The kanamycin resistant recombinant clones were pooled and cosmid DNA was prepared using standard plasmid isolation methods. The variability of members of the library was established.

A tuberculosis complex strain of lowered virulence for guinea pigs (referred to subsequently as avirulent) was cultured in roller bottles and organisms

were prepared and electroporated with a library of pYUB178::virulent-tuberculosis-complex-DNA. The electroporated organisms were plated onto media containing kanamycin and kanamycin resistant clones were pooled to form a library. Each member of this library had the chromosome of the avirulent tuberculosis organism into which a cosmid with an insert of genomic DNA from a virulent tuberculosis complex strain was integrated. The library was cultured in liquid media and aliquots were inoculated into guinea pigs. Separate guinea pigs were also inoculated with the matching avirulent tuberculosis complex strain as a control. The most clear cut distinction between virulent and avirulent strains was in the presence or absence of gross lesions in the spleen.

The method for virulence testing in guinea pigs was adapted from the procedures described in the Trudeau Mycobacterial Culture Collection catalogue, (Anon, 1972). Albino, outbred guinea pigs were inoculated subcutaneously in the flank. Libraries and individual strains of mycobacteria were inoculated into at least three guinea pigs which were kept in filtered-air, ventilated animal cages. Animals were sacrificed approximately 6 and 13 weeks after inoculation and examined for the presence of gross lesions of tuberculosis. Samples from the injection site, the prefemoral lymph nodes and spleen were cultured for mycobacteria using previously described methods. Formalin-fixed tissues, from the spleen, liver, kidney and lung were embedded in paraffin, sectioned at 3-5 μm, and stained with either hematoxylin and eosin (HE) or by the Ziehl-Neelsen method.

A. Virulent Tuberculosis Strain Used to Make Cosmid Library

A virulent *M. bovis* strain was isolated from bovine tissue submitted to the Wallaceville Animal

Research Centre, Upper Hutt, New Zealand. The strain, isolated from bovine tissue with the accession number 89/5276, was designated WAg200 and was cultured as described previously (Collins and de Lisle 1984). The strain was also shown to be virulent for guinea pigs. Bacteriological identification of the strain as *M. bovis* was based on colony morphology, slow growth, acid-fast staining, susceptibility to thiophene-2-carboxylic acid hydrazide and isoniazid, and growth on pyruvate-supplemented but not glycerol-supplemented media. The strain was also characterized by restriction fragment analysis (Collins et al. 1993). In infected animal experiments described below, bacteriological identification of reisolated *M. bovis* strains was based on colony morphology, slow growth and growth on pyruvate-supplemented media.

B. DNA Preparation

*M. bovis* WAg200 was cultured under biosafety containment at 37°C on 40 x 85 mm petri dishes of 7H11 Middlebrook (Difco) media containing oleic acid, albumin, dextrose, serum, lysed red blood cells, 0.05% polyoxyethylene sorbitan monooleate (Tween-80) and pyruvate (Gallagher and Norwill 1977). The organisms were harvested into 7 Falcon tubes each containing 50 ml phosphate buffered saline (0.14 M NaCl, 4 mM KCl, 8 mM $Na_2HPO_4$, 2 mM $KH_2PO_4$; pH 6.5) and inactivated by heating at 75°C for 35 min. After centrifugation, the yield in each tube was 1-1.5 g wet weight organisms. Genomic DNA was prepared from the organisms using a scaled up version of the method described by van Soolingen et al. (1991). The total yield of DNA after extraction of all organisms was 300 µg in 1 ml.

C. *E. coli* cosmid library of *M. bovis* WAg200

*M. bovis* WAg200 DNA was partially digested with a range of concentrations of Sau3AI and digestions having the largest yield of 30-50 kb fragments were selected after analytical electrophoresis on 0.4% agarose gels (Jacobs et al. 1991). Fragments of 30-50 kb from these digestions were prepared using sucrose gradient centrifugation (Weis 1991) and ligated to BclI-digested pYUB178 DNA that had been treated with calf intestinal phosphatase. The final genomic DNA concentration in the 10 µl ligation mixture was 200 ng/µl and the DNA molar ratio of insert to vector was 1:20. Four µl of the ligation mixture was *in vitro*-packaged with the GigaPack II Gold Packaging Extract (Stratagene, La Jolla, CA) according to the manufacturer's procedure. The *in vitro*-packaged lysate was transduced, using previously described methods (Jacobs et al. 1991), into *E. coli*. The kanamycin resistant recombinant clones were pooled and inoculated into LB broth containing 25 µg/ml kanamycin. Cosmid DNA was prepared using standard plasmid isolation methods of alkaline lysis and cesium chloride gradient centrifugation (Sambrook et al. 1989). Aliquots of the library were stored frozen at -70°C and cosmid DNA preparations were stored at -20°C.

The total number of recombinant *E. coli* clones produced was approximately 20,000. These clones were pooled and the library of pYUB178::*M. bovis* WAg200 cosmids was amplified as a plasmid preparation. This preparation was performed by culturing the pooled clones in 750 ml LB media containing 25 µg/ml kanamycin. Both before and after the plasmid amplification of the library, cosmids from 20 randomly selected clones were shown to have different restriction patterns.

Plasmids and *M. bovis* strains used in this study are listed in Tables 1 and 2.

TABLE 1
*M. bovis* strains used in this study

| *M. bovis* strain or clone | Description | Source |
|---|---|---|
| ATCC35721 | Low virulence strain | ATCC |
| WAg200 | Virulent strain isolated in N.Z. | G. de Lisle |
| WAg300 | ATCC35721 containing pUHA1 | This study |
| WAg301 | ATCC35721 containing pUHA3 | This study |
| WAg302 | ATCC35721 containing pUHA4 | This study |
| WAg303 | ATCC35721 containing pUHA5 (Junction Fragment Pattern 1) | This study |
| WAg304 | ATCC35721 containing pUHA5 (Junction Fragment Pattern 2) | This study |
| WAg305 | ATCC35721 containing pUHA5 (Junction Fragment Pattern 3) | This study |
| WAg306 | ATCC35721 containing pUHA6 (Junction Fragment Pattern 1) | This study |
| WAg307 | ATCC35721 containing pUHA6 (Junction Fragment Pattern 2) | This study |
| WAg308 | ATCC35721 containing pUHA6 (Junction Fragment Pattern 3) | This study |
| WAg309 | ATCC35721 containing pUHA7 (Junction Fragment Pattern 1) | This study |
| WAg310 | ATCC35721 containing pUHA7 (Junction Fragment Pattern 2) | This study |
| WAg311 | ATCC35721 containing pUHA7 (Junction Fragment Pattern 3) | This study |

| | | |
|---|---|---|
| WAg320 | ATCC35721 containing 3 kb fragment of WAg200 that restores virulence | This study |

TABLE 2
Plasmids used in this study

| Plasmid | Description | Source |
|---|---|---|
| pYUB178 | Integrating cosmid shuttle vector | W. Jacobs |
| pUHA1 | pYUB178::WAg200 cosmid which restores virulence to ATCC35721 | This study |
| pUHA2 | pYUB178 containing 6 kb of pUHA1 insert | This study |
| pUHA3 | pYUB178::WAg200 cosmid overlapping pUHA2 | This study |
| pUHA4 | pYUB178::WAg200 cosmid overlapping pUHA2 | This study |
| pUHA5 | pYUB178::WAg200 cosmid overlapping pUHA2 | This study |
| pUHA6 | pYUB178::WAg200 cosmid overlapping pUHA2 | This study |
| pUHA7 | pYUB178::WAg200 cosmid overlapping pUHA2 | This study |
| pUHA8 | pYUB178 with *PacI* sites on both sides of the *BclI* cloning site | This study |
| pUHA9 | pBluescript II KS(+) with *PacI* sites on both sides of the *BclI* site | This study |

| | | |
|---|---|---|
| pUHA11 | pUHA9 containing 3 kb fragment from WAg320 | This study |
| pUHA16 | pUHA11 with 3 kb fragment in reverse orientation | This study |

D. Transformation of cosmid library into avirulent *M. bovis*

The receptor strain used was *M. bovis* ATCC35721 which had lowered virulence for guinea pigs. For simplicity this strain is subsequently referred to as avirulent. It was inoculated into 2 x 100 ml Middlebrook 7H9 broth (Difco) containing albumin, glucose, glycerol and Tween-80 as described (Jacobs et al. 1991). The cultures were grown in roller bottles at 1 revolution/min to an O.D. at 600nm of 0.18. The organisms were washed and concentrated to a volume of 1 ml in cold 10% glycerol and 0.4 ml were electroporated with 4 µl of pYUB178::*M. bovis* WAg200 cosmid library DNA (1 µg/µl) as described by Jacobs et al. (1991). After electroporation, the organisms were cultured at 37°C on the same media used for DNA preparation but without the addition of oleic acid, serum or lysed red blood cells and with the addition of 1% sodium pyruvate and 10 µg/ml kanamycin. Approximately 4000 clones of *M. bovis* ATCC35721 (pYUB178::*M. bovis* WAg200) were obtained and pooled. A control electroporation of 400 µl organisms without added plasmid DNA yielded no kanamycin resistant colonies. Fifteen *M. bovis* ATCC35721 (pYUB178::*M. bovis* WAg200) clones were selected before pooling and subcultured for DNA preparation in 3-5 ml of the same media used for culturing *M. bovis* ATCC35721. Genomic DNA of recombinants, extracted by the method of van Soolingen et al. (1991), was characterized by restriction fragment digestion with *Pst*I, electrophoresis, Southern blotting and hybridization with a probe of pYUB178. This revealed

the junction fragments of the integrated cosmid and is referred to below as junction fragment analysis. In all cases the fragment patterns were different.

E. Protocol to assess virulence of tuberculosis complex strains

The method for virulence testing in guinea pigs was adapted from the procedures described in the Trudeau Mycobacterial Culture Collection catalogue, (Anon, 1972). Albino, outbred guinea pigs were inoculated subcutaneously in the flank. Libraries and individual strains of mycobacteria were inoculated into guinea pigs which were kept in filtered-air, ventilated animal cages. Animals were sacrificed approximately 6 and 13 weeks after inoculation and examined for the presence of gross lesions of tuberculosis. Samples from the injection site, the prefemoral lymph nodes and spleen were cultured for mycobacteria using previously described methods (Collins and de Lisle 1984). Formalin-fixed tissues, from the spleen, liver, kidney and lung were embedded in paraffin, sectioned at 3-5 μm, and stained with either hematoxylin and eosin (HE) or by the Ziehl-Neelsen method.

i. First inoculation experiments in guinea pigs

The level of virulence in guinea pigs of *M. bovis* ATCC35721 was assessed by the sites in which gross lesions were found (Table 3). There were no such lesions in the spleen. This indicated that *M. bovis* ATCC35721 was not sufficiently virulent to pass through the lymph nodes draining the injection site and enter the systemic circulation in sufficient numbers to cause gross lesions to occur in the spleen.

TABLE 3

Gross lesions in animals sacrificed 92 days after infection with a 0.2 ml inoculum of *M. bovis* ATCC35721 containing $1.9 \times 10^7$ colony forming units (CFU).

| Guinea pig | Injection site | Prefemoral lymph nodes | Spleen |
|---|---|---|---|
| A | + | + | - |
| B | + | + | - |
| C | + | + | - |

In a subsequent experiment, the virulence of the *M. bovis* ATCC35721(pYUB178::*M. bovis* WAg200) library was assessed at two time intervals and gross lesions were identified as shown in Tables 4 and 5.

TABLE 4

Gross lesions in animals sacrificed 50 days after infection with a 0.2 ml inoculum of *M. bovis* ATCC35721(pYUB178::*M. bovis* WAg200) library containing approximately $10^6$CFU.

| Guinea pigs | Injection site | Prefemoral lymph nodes | Spleen |
|---|---|---|---|
| A | +/- | + | - |
| B | + | + | - |
| C | + | + | + |

TABLE 5

Gross lesions in animals sacrificed 89 days after infection with a 0.2 ml of inoculum of *M. bovis* ATCC35721(pYUB178::*M. bovis* WAg200) library containing approximately $10^6$ CFU.

| Guinea pigs | Injection site | Prefemoral lymph nodes | Spleen |
|---|---|---|---|
| A | + | + | + |
| B | + | + | + |
| C | + | + | + |

ii. Characterization of recombinant *M. bovis* from guinea pigs

Prefemoral lymph node and spleen tissues of all guinea pigs were cultured for the presence of *M. bovis*. Apart from spleen tissue from guinea pig A in the 50 day group, *M. bovis* organisms were isolated from all these tissues. Over 160 individual clones representing all lesion-containing prefemoral lymph nodes and spleens were subcultured and their genomic DNA subjected to junction fragment analysis. Approximately 80% of all clones had the same junction fragment pattern. Clones which gave this pattern were found in all *M. bovis* containing tissues. One of these ATCC35721(pYUB178::*M. bovis* WAg200) clones containing the predominant junction fragment pattern designated as WAg300 was used for further experiments below.

iii. Second inoculation experiment in guinea pigs

In this experiment the virulence of *M. bovis* WAg300 and *M. bovis* ATCC35721 were compared concurrently. Results are given in Tables 6 and 7.